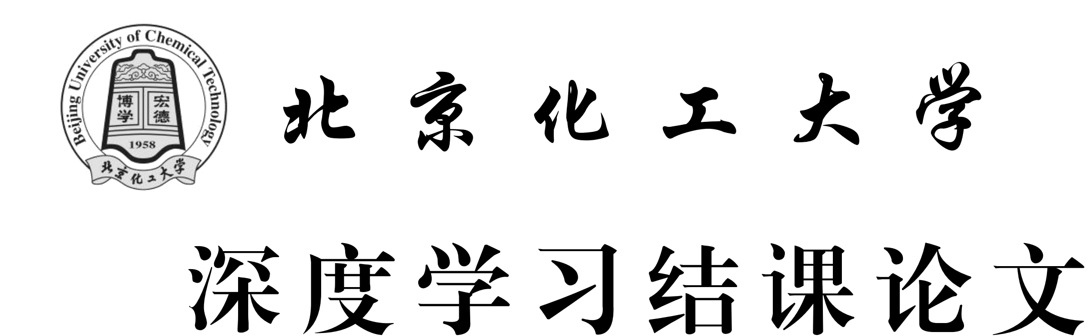
班　　级：信研 2308
学　　号：2023200841



北 京 化 工 大 学

深度学习结课论文

题　目　　从自然场景图像中学习可动画的三维人脸模型

专　　业　信息与通信工程

学　　生　蒋翠莲

任课教师　李瑞瑞

2024 年 1 月 3 日

从自然场景图像中学习可动画的三维人脸模型

摘 要

虽然目前的基于单幅图像的三维人脸重建方法可以恢复精细的几何细节，但是这些方法具有局限性。某些方法生成的人脸无法真实地设置动画，因为它们没有对皱纹如何随表情变化进行建模。其他方法是在高质量的面部扫描上训练的，并且不能很好地推广到自然场景的图像。报告中使用的方法，能回归三维人脸形状和动画的细节，这些细节是特定于个人的但能随着表情的变化。该方法的模型经过训练，可以从由特定于人的细节参数和通用表情参数组成的低维潜在表示中鲁棒地生成UV位移图，而回归量则经过训练，可以从单个图像中预测细节，形状，表情，姿势和照明参数。为了实现这一点，该方法引入了一种新的细节一致性损失，从依赖于表情的皱纹中分离出特定于人的细节。这种解缠使得能够通过控制表情参数，同时保持个人特定细节不变，来合成逼真的个人特定皱纹。该方法是从自然场景得图像中学习的，没有配对的三维数据监督。

关键词：三维人脸重建，深度学习，人脸细节

Learning Animable 3D Face Model from Natural Scene Images

Abstract

Although the current 3D face reconstruction methods based on a single image can recover fine geometric details, these methods have limitations. The faces generated by some methods can't be really animated because they don't model how wrinkles change with expressions. Other methods are trained on high-quality facial scanning, and cannot be well extended to images of natural scenes. The method used in the report can return to the details of three-dimensional face shapes and animations, which are specific to individuals but can change with expressions. The model of this method can be trained to generate a UV displacement map from a low-dimensional potential representation composed of person-specific detail parameters and general expression parameters, while the regression quantity can be trained to predict details, shapes, expressions, postures and lighting parameters from a single image. In order to achieve this, this method introduces a new loss of detail consistency, which separates people-specific details from wrinkles that depend on expressions. This unwrapping makes it possible to synthesize realistic personal specific wrinkles by controlling expression parameters while keeping personal specific details unchanged. This method is learned from images of natural scenes, and there is no paired 3D data supervision.

Keywords: 3D face reconstruction, deep learning, face details.

目录

第 1 章 绪论

1.1 研究背景

人脸作为人类最具辨识度的生物特征，在日常生活中传递了大量的信息，包括身份情绪和年龄等，这些信息具有直观性、唯一性等特点。因此人脸研究具有颇多的价值，因此成为了计算机视觉领域的热门研究方向。与二维人脸图像不同，三维人脸能够在空间中多角度的展示人脸，因此其能将人脸的形状、姿态和纹理等信息更加丰富的表现出来，且视角转换以及角度遮挡等情况不会对人脸在三维空间中的表征产生影响，模型的鲁棒性大大提高。但是，由于人脸携带了身份、性别、种群、年龄和情感等大量信息，因此要在三维空间中将这些蕴含的人脸信息充分、准确的表现出来就需要重建出高质量的三维人脸模型。如何重建出高质量的三维人脸模型仍是三维人脸重建技术中具有挑战性的问题。

三维人脸重建相关技术被广泛应用于多个领域，主要包括以下场景:

(1)智慧医疗领域。三维人脸重建技术在面部整形方面有着广泛的应用。在手术前,医生可以使用三维人脸重建软件对病人的脸部进行三维建模，从而加深对病人情况的认知。此外，也可以通过对重建的三维人脸模型进行美容、化妆、编辑等操作，将手术后的人脸结果展示给患者，从而使医生和病人间的沟通交流更加方便。

(2)影视娱乐领域。随着影视观众对影像视听品质以及游戏用户对多元化游戏体验需求的提升，过去的二维人脸图像处理技术已经很难满足用户日益增长的需求因此三维人脸重建技术在电影、电视与游戏领域得到了广泛的应用。迄今为止，已经有多部电影将经过三维重建的人脸模

型呈现给观众，诸如国产动漫《大圣归来》和《哪吒之魔童降世》等，很好的将传统中国文化中的人物形象与三维人脸重建技术相结合，得到了口碑与票房双丰收。在游戏领域三维人脸重建技术应用的时间可能更早，早在2004年就有了中国的第一款3D网游《天之游侠》，并得到了玩家们的极高评价，此后越来越多的游戏公司开始研发能给用户带来更好体验感与真实感的三维游戏，在开发过程中也同时引入了三维人脸重建技术如今一些游戏不仅可以提供对三维人脸模型的捏脸服务,甚至可以由玩家提供自己的人脸照片实现个性化的人脸定制，增强了可玩性以及用户的参与感。

(3)人脸识别领域。每个人的脸部特征都具有唯一性，即使存在长相相似的人但他们的脸部特征也不会完全相同，因此人脸可以如同指纹一样作为唯一特征来区分不同人的身份。科研工作者们据此研究出了许多人脸识别的算法。随着深度学习方法的兴起，人脸识别技术在金融、安全检查、打卡考勤等多个领域取得了广泛的应用。然而，当前人脸识别相关技术大多数是使用二维人脸图像来进行识别，只有极少数方法使用三维人脸技术，这就导致了在大姿态以及极端光照情况下，难以实现有效的人脸识别，无法满足实际应用需求。甚至有人使用二维人脸图像冒充真实三维人脸来欺骗人脸识别设备。与二维人脸图像不同的是三维人脸模型具有更加丰富的深度与纹理信息且不受观察视角的影响，因此将三维人脸重建技术与人脸识别相结合将更好的提高其识别准确度。

由此可见，作为计算机视觉最热门的研究方向之一，三维人脸重建技术有着许多的实际应用场景，并且不断给人类带来更好的生活体验。在当下工业界为了获得三维人脸模型仍然采用结构光相机以或三维激光扫描仪来采集人脸形状与纹理信息。使用这类设备采集的结果进行重建得到的三维人脸模型精度很高，形状也十分逼真，但昂贵的设备成本

以及复杂耗时的过程使得该方法门槛较高，在实际应用中受到了很大的限制。二维人脸图像相较于三维人脸数据获取门槛低且日常生活中使用手机、相机等设备就能轻松获得，因此基于单张二维人脸图像进行三维人脸重建成为了学者们研究的重点方向。伴随着深度学习技术的迅猛发展，三维人脸重建领域也迎来了快速发展，这对于未来的人类生活发展有着重大的研究价值和实际应用意义。

1.2 国内外研究现状

三维人脸重建技术旨在从二维人脸图像中获得准确的三维人脸数据，并通过获取的三维人脸数据重构出三维人脸模型，是目前国内外学者和工业界关注的热点。早期，研究人员使用简单的建模方法和模板变形来重建出三维人脸:后来，在 3DMM 人脸可形变模型被提出后，三维人脸重建技术围绕该模型进行不断完善发展:现在，随着深度学习的快速发展，使用深度学习方法实现端到端的三维人脸重建成为了行业主流。

该报告主要研究的是基于深度学习的单张人脸图像的三维人脸重建的方法。

近年来，随着深度学习技术的发展，计算机视觉的相关领域也因此得到了快速发展由于无法捕捉到图像的深层特征导致人脸模型的表达能力受限，传统的三维人脸重建方法已经难以满足日益增长的精细化三维人脸模型需求。同时传统的三维人脸重建方法中间过程较多，需要经过繁琐的操作才能得到重建的三维人脸模型。而深度学习模型则可以根据学习任务选择损失函数进行约束，从而实现对输入图像的自适应深层特征提取，这有效的弥补了传统三维人脸重建方法中模型表达能力的不足。同时深度学习方法中端到端的设计也简化了搭建模型的中间过程，弥补了传统三维人脸重建方法中的不足和缺陷。由于以上的优点，使用深度学习方法的三维人脸重建技术成为了该领域的新研究热门。

基于深度学习的三维人脸重建方法按是否需要三维人脸数据可以分为有监督方法和无监督方法。有监督方法需要对应的三维人脸数据来优化训练人脸重建网络。一些方法[2][3]将卷积神经网络(CNN)引入到 3DMM 模型中，使用卷积神经网络对 3DMM 模型的参数进行直接预测。DOU[3]等人提出了一种在卷积神经网络的不同层上对 3DMM 模型参数进行预测的方法，该方法结合了卷积神经网络与 3DMM 模型，并在网

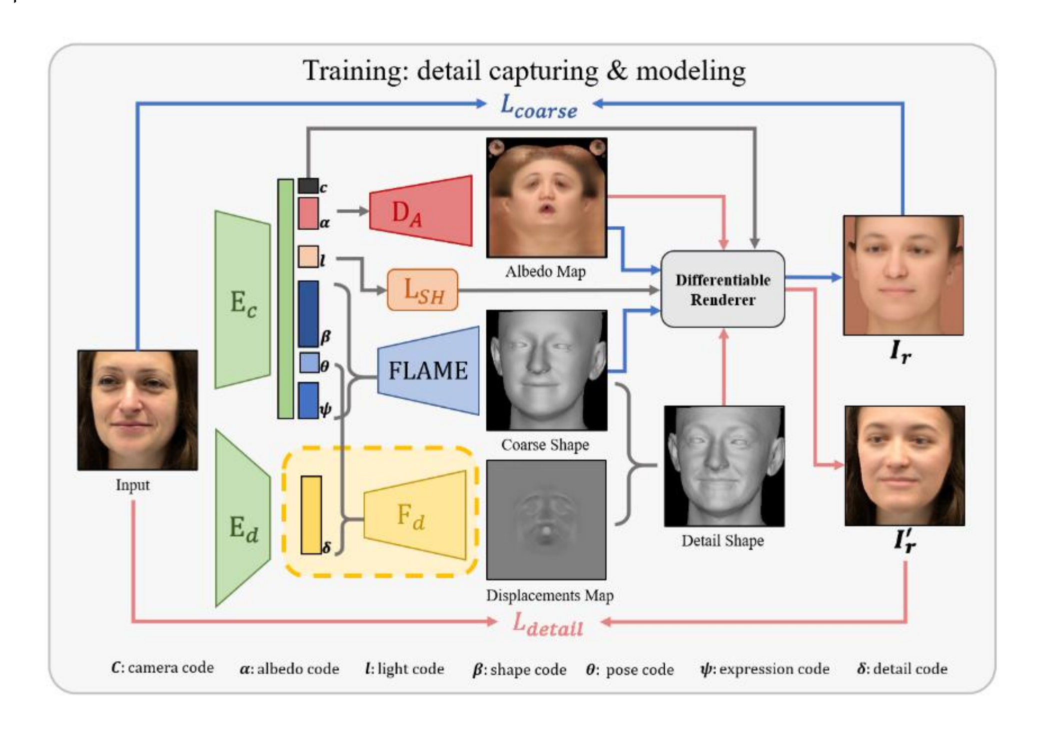
络的不同层上对表情、身份等参数分别进行预测。Richardson[2]等人提出了一种逐步精细化的三维人脸重建算法，该算法由 CoarseNet 和 FineNet 两个网络模块组成，先由 CoarseNet 模块通过 3DMM 方法恢复出粗糙人脸形状，再由 FineNet 模块对粗人脸形状进行不断地细化。Zhu 等人[1]提出了基于级联卷积神经网络的三维人脸重建方法，通过级联卷积神经网络将 3DMM 模型与输入的二维人脸图像拟合，进行密集人脸重建与对齐，展现出了优良的人脸重建与对齐效果。VRNet[5]将人脸看成是从耳后平面到鼻尖平面的 200 个横切片，使用卷积神经网络直接回归得到每个切片进而重建三维人脸模型，然而该方法会带来人脸对齐以及重建分辨率不易扩展等问题。PRNet[6]提出了一种基于 UV 位置映射的 2D 表示，它能够在 UV 空间中记录单张人脸图像的 3D 形状信息，通过训练卷积神经网络重建出三维人脸形状并与 UV 空间的 3D 纹理对齐。

尽管基于深度学习的三维人脸重建方法已经获得了快速发展，但由于缺少三维人脸数据和相应的三维标注信息，三维人脸重建模型的性能受到了严重的限制。为此，科研工作者们提出了无监督的三维人脸重建方法，从而缓解了三维人脸数据标签不足的问题无监督方法则是将生成的三维人脸模型渲染到二维，通过与输入图像的不断拟合来优化人脸重建网络。Pan[7]等人利用 GAN 网络从二维图像中挖掘 3D 几何线索，从而实现以无监督的方式从单张二维人脸图像中恢复出三维人脸形状。Shang[8]等人使用单张人脸图像生成多个视角下的人脸图片，并利用同一人脸在不同视角下的一致性约束来增强三维人脸模型的表达能力。Thewlis[9]等人使用等方差来学习密集地标，并使用学习到的地标信息来恢复 2D 对象对应的 3D 几何形状。Li[10]等人通过利用三维网格和深度图像的致性，以无监督的方式来拟合优化网络，可以实现在光照不足的

情况下进行三维人脸的重建。DEA[11]通过进一步分解图像的反照率和阴影，并约束带有小瓶颈嵌入的自动编码器来预测面部三维形状。Zhang[12]等人提出了一种新颖的学习聚合和个性化框架，解决了无监督三维人脸重建在大姿态和更高分辨率重建效果不佳的问题。当下流行的方法在这些工作的基础上有了较大的突破，Wu[13]基于无监督的方法提出将二维人脸图片通过对称性解耦为视角，光照，反照率，深度等因素并将他们组合起来得到重建的三维人脸模型。

第 2 章 本报告的关键技术

本报告的理论和代码基础来源于论文：Learning an Animatable Detailed 3D Face Model from In-The-Wild Images。提出了两阶段人脸重建的方法，第一阶段先重建出人脸的粗略形状，第二阶段恢复出人脸细节。



图一 两阶段人脸重建的方法

如图一所示，在第一阶段的粗略重建中，先是将一张人脸图像出入到一个编码器中（原文中是 ResNet50），得到 FLAME 人脸模型的形状系数，表情系数，纹理系数，姿势，光照，相机参数，反射率等参数。再利用这些参数通过一个可微分渲染器得到一张二维图像。最后对渲染得到的二维图像和原始的输入图像作损失并计算损失函数，更新网络参

数。将粗略重建的到的形状系数，表情系数，姿势系数送入 FLAME 解码器，就能得到三维人脸的顶点，也就是三维人脸。

在粗略重建训练完成后，将其网络参数固定，开始进行细节重建。细节重建的输入也是二维图像，通过一个编码器（原文中是 ResNet50）将输入图像编码为 128 维的细节编码，再将细节编码与粗略重建得到的表情参数和下巴姿势参数结合，送入到解码器中，得到一张置换贴图，再将该置换贴图应用于粗略重建的结果，就得到了最终的结果。

使用的三维人脸模型是 FLAME，这是一个基于统计学的三维人脸模型，具有 5023 个顶点，能够使用形状系数，表情系数，姿势参数来生成不同的三维人脸。

2.1 损失函数

总的损失函数表示为：

$$L_{coarse} = L_{lmk} + L_{eye} + L_{pho} + L_{id} + L_{sc} + L_{reg}$$

人脸关键点损失表示为：

$$L_{lmk} = \sum_{i=1}^{68} \|\mathbf{k}_i - s\Pi(M_i) + \mathrm{t}\|_1$$

计算的是输入人脸图像的 68 个关键点与生成的 flame 人脸模型的 68 个关键点投影到图像平面后的对应点之间的距离。

闭眼损失表示为：

$$L_{eye} = \sum_{(i,j)\in E} \left\|\mathbf{k}_i - \mathbf{k}_j - s\Pi(M_i - M_j)\right\|_1$$

与人脸关键点损失类似，闭眼损失计算的是眼部关键点之间的相对偏移。

光度损失：

$$L_{pho} = \|V_I \odot (I - I_r)\|_{1,1}$$

I 是原始的输入图像，Ir 是渲染后得到的图像，VI 是对应输入图像的 mask，表示人脸的可见区域。

身份一致性损失：

$$L_{id} = 1 - \frac{f(I)f(I_r)}{\|f(I)\|_2 \cdot \|f(I_r)\|_2}$$

f（I）是使用预训练的人脸识别网络对原始输入人脸图像提取的特征，f（I_r）是对渲染后得到的图像提取的特征。形状一致性损失：

$$L_{sc} = L_{coarse}(I_i, \mathcal{R}(M(\boldsymbol{\beta}_j, \boldsymbol{\theta}_i, \boldsymbol{\psi}_i), B(\boldsymbol{\alpha}_i, \mathbf{l}_i, N_{uv,i}), \mathbf{c}_i))$$

对于同一个人的不同照片，其形状系数应保持不变，再固定形状系数后，其他系数保持不变，计算上述所有损失值。

2.2 细节重建过程的损失函数

总的损失函数表示为：

$$L_{detail} = L_{phoD} + L_{mrf} + L_{sym} + L_{dc} + L_{regD}$$

细节光照度损失函数：

$$L_{pho} = \|V_I \odot (I - I_r)\|_{1,1}$$

与粗略重建过程中的光照度损失类似，这里的 Ir 是渲染图像应用置换贴图后的图像。

隐式多样化马尔可夫随机场（ID-MRF）损失：

$$L_{mrf} = 2L_M(conv4_2) + L_M(conv3_2)$$

在 VGG19 的第 3_2 层和第 4_2 层上计算损失。

软对称性损失：

$$L_{sym} = \|V_{uv} \odot (D - flip(D))\|_{1,1}$$

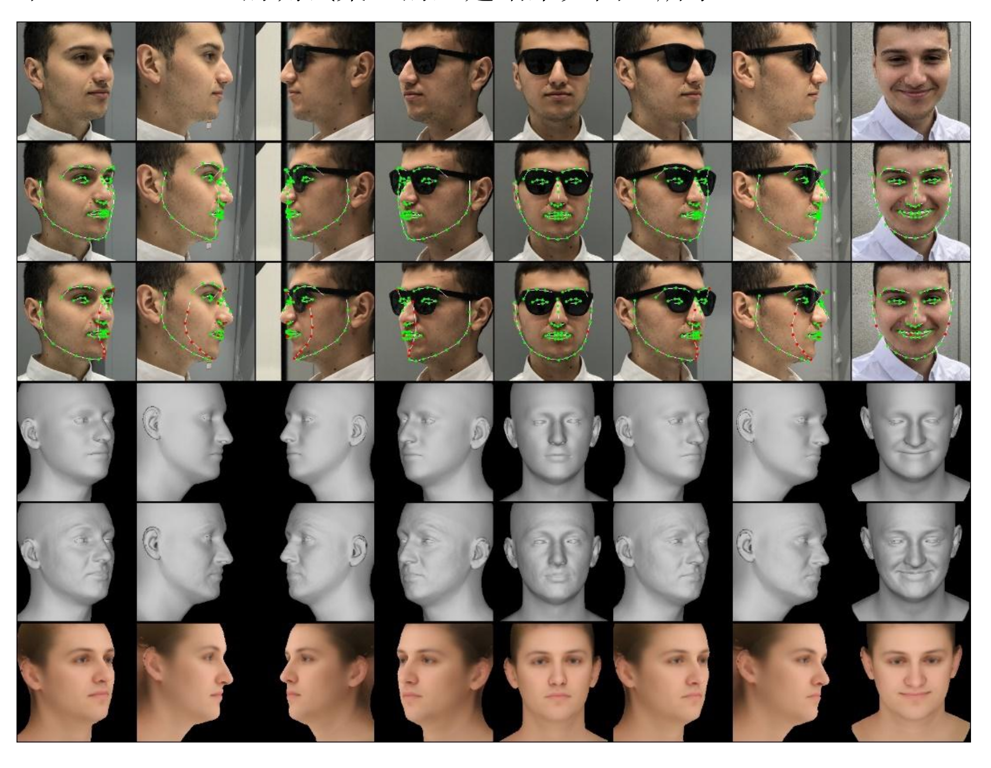
为了增加自遮挡的鲁棒性，添加了一个软对称损失来正则化不可见的面部部分。Vuv 表示 UV 空间中的面部皮肤掩码，flip 表示水平翻转操作，D 是由置换贴图得到的法线贴图。

第 3 章 实验结果

在该论文开源代码的基础上，使用收集到的人脸数据集进行了重新训练和评估训练结果。

数据集组成：VggFace2，CelebA，AFLW，再将这些数据集按 8：2 的比例划分为训练集和验证集。其他参数均按照原论文设置，其中 batch size 设置为 4。

在 Now dataset 的测试集上的重建结果如图二所示。



图二 重建结果

第一行为原始的输入图像；第二行为预测出的 68 个二维人脸关键点；第三行为预测出的 68 个三维人脸关键点，其中绿色的点为可见点，

红色的点为不可见点；第四行为重建出的粗略的人脸形状；第五行为添加了置换贴图的人脸形状；第六行为添加了纹理贴图的人脸形状。可以看出，重建的效果较好，对于大姿态和有遮挡的情况也能得到人脸模型。

在 Now dataset 的验证集上（其官网地址为：https://now.is.tue.mpg.de/）进行定量评估，评估方式为计算预测的人脸顶点与真实的人脸顶点之间的距离，并计算所有距离的平均值，中位数以及标准差。根据官网给出的代码以及验证集，计算得到的重建误差为：中位数：1.18mm，平均值：1.46mm，，标准差：1.25mm。Now dataset 测试集重建误差排行版如图三所示。

NoW Home Dataset Non-Metrical Evaluation Metrical Evaluation Downloads License Sign In

Rank	Method	Median (mm)	Mean (mm)	Std (mm)	Error files	PDF	Code
1.	TokenFace [Zhang et al., ICCV 2023]	**0.76**	**0.95**	**0.82**	Download (415 MB)	PDF	
2.	MICA [Zielonka et al., ECCV 2022]	0.90	1.11	0.92	Download (415 MB)	PDF	Code
3.	AlbedoGAN [Rai et al., 2023]	0.98	1.21	0.99	Download (415 MB)	PDF	Code
4.	Wood et al. [ECCV 2022]	1.02	1.28	1.08	Download (415 MB)	PDF	
5.	FOCUS [Li et al. 2022]	1.04	1.30	1.10	Download (415 MB)	PDF	Code
6.	CCFace [Yang et al., IEEE TMM 2023]	1.08	1.35	1.14	Download (415 MB)	PDF	
7.	DECA [Feng et al., SIGGRAPH 2021]	1.09	1.38	1.18	Download (415 MB)	PDF	Code
8.	Deep3DFaceRecon PyTorch [Deng et al., CVPRW 2019]	1.11	1.41	1.21	Download (415 MB)	PDF	Code
9.	PyMAF-X [Zhang et al. 2022]	1.13	1.42	1.20	Download (415 MB)	PDF	Code
10.	RingNet [Sanyal et al., CVPR 2019]	1.21	1.53	1.31	Download (415 MB)	PDF	Code

图三 重建误差排行版

第 4 章 结论

该论文提出的方法通过从自然场景下的人脸图像数据集中学习一个可动画的细节模型来从单张图像中重建出具有丰富表情和可动画细节的人脸模型。该方法是在没有 2 维到 3 维监督的自然场景下的数据集上训练的。该方法提出的细节重建能够将特定于人的细节和特定于表情的细节解缠，这使得能生成可动画的人脸细节，能应用于动画，形状改变，皱纹转换等方面。同时由于该方法的准确性，可靠性以及速度，使得该方法能应用于人脸再现，虚拟数字人的创造等方面。

该论文提出的人脸重建方法对于遮挡以及大范围的表情具有良好鲁棒性，但是在大姿态的情况下可能会失败。此外，该方法能从单张人脸图像中恢复出特定于人的人脸细节，同时这些细节能随着表情的变化而变化。同时，该方法完全开源，对于三维人脸重建的研究工作具有重大指导作用。

参考文献

[1] Zhu X, Yang F, Huang D, et al. Beyond 3dmm space: Towards fine-grained 3d face reconstruction[C]//Computer Vision-ECCV 2020: 16th European Conference, GlasgowUK, August 23-28,2020, Proceedings, Part VII 16. Springer International Publishing.2020:343-358.

[2] Richardson E, Sela M, Or-El R, et al. Learning detailed face reconstruction from a single image[C]//Proceedings of the IEEE conference on computer vision and pattern recognition.2017:1259-1268.

[3] Dou P, Shah S K, Kakadiaris I A. End-to-end 3D face reconstruction with deep neural networks[C]//proceedings of the IEEE conference on computer vision and patternrecognition.2017:5908-5917.

[4] Zhu X, Liu X, Lei Z, et al. Face alignment in full pose range: A 3d total solution[J]. IEEEtransactions on pattern analysis and machine intelligence, 2017,41(1): 78-92.

[5] Jackson A S, Bulat A, Argyriou V, et al. Large pose 3D face reconstruction from a single image via direct volumetric CNN regression[C]//Proceedings of the IEEE international conference on computer vision.2017: 1031-1039

[6] Feng Y, Wu F, Shao X, et al. Joint 3d face reconstruction and dense alignment with position map regression network[C]/Proceedings of the European conference on computer vision(ECCV).2018:534-551.

[7] Pan X, Dai B, Liu Z, et al. Do 2d gans know 3d shape? unsupervised 3d shape reconstruction from 2d image gans[J]. arXiv preprint arXiv:2011.00844, 2020.
[8] Shang J, Shen T, Li S,et al. Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry consistency[C]//Computer Vision-ECCV 2020: 16thEuropean Conference, Glasgow, UK,August 23-28, 2020, Proceedings, Part XV. Cham:Springer International Publishing, 2020: 53-70.
[9] Thewlis J, Bilen H, Vedaldi A. Modelling and unsupervised learning of symmetric deformable object categories[J]. Advances in Neural Information Processing Systems,2018,31.
[10] Li P, Pei Y, Zhong Y, et al. An unsupervised approach for 3D face reconstruction from a single depth image[C]//Advances in Computer Graphics: 37th Computer Graphics International Conference, CGI 2020, Geneva， Switzerland, October 20-23， 2020.Proceedings 37.Springer International Publishing, 2020: 206-219.
[11] Shu Z, Sahasrabudhe M, Guler R A, et al. Deforming autoencoders: Unsupervised disentangling of shape and appearance[C]//Proceedings of the European conference on computer vision(ECCV).2018:650-665.
[12] Zhang Z, Ge Y, Chen R, et al. Learning to aggregate and personalize 3d face from in-the-wild photo collection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.2021:14214-14224.

[13] Wu S, Rupprecht C, Vedaldi A. Unsupervised learning of probably symmetric deformable3d objects from images in the wild[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.2020:1-10